
DATA-DRIVEN COLLECTIVE VARIABLES FOR ENHANCED SAMPLING

A PREPRINT

Luigi Bonati

Department of Physics, ETH Zurich, 8092 Zurich, Switzerland
and Institute of Computational Sciences, Università della Svizzera italiana (USI),
via Buffi 13, 6900 Lugano, Switzerland

Valerio Rizzi

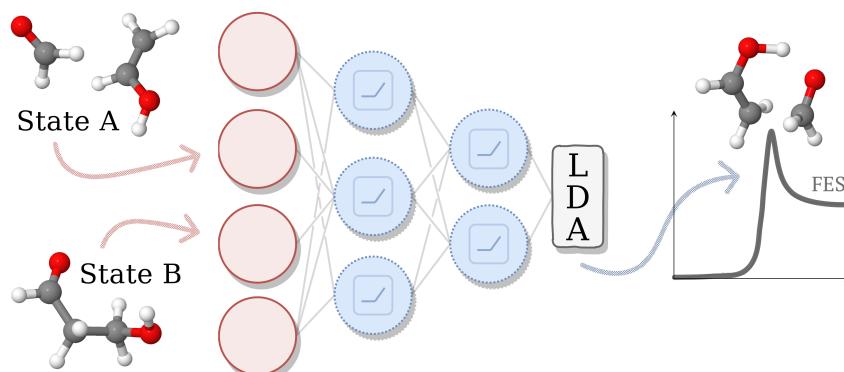
Department of Chemistry and Applied Biosciences, ETH Zurich, 8092 Zurich, Switzerland
and Institute of Computational Sciences, USI, via Buffi 13, 6900 Lugano, Switzerland

Michele Parrinello

Department of Chemistry and Applied Biosciences, ETH Zurich, 8092 Zurich, Switzerland,
Institute of Computational Sciences, USI, via Buffi 13, 6900 Lugano, Switzerland,
and Italian Institute of Technology, Via Morego 30, 16163 Genova, Italy

ABSTRACT

Designing an appropriate set of collective variables is crucial to the success of several enhanced sampling methods. Here we focus on how to obtain such variables from information limited to the metastable states. We characterize the states by a large set of descriptors and employ neural networks to compress this information on a lower-dimensional space, using Fisher's linear discriminant as objective function to maximize the discriminative power of the network. We benchmark this method on alanine dipeptide, using the non-linearly separable dataset composed by atomic distances. We then study an intermolecular aldol reaction characterized by a concerted mechanism. The resulting variables are able to promote sampling by drawing non-linear paths in the physical space connecting the fluctuations between metastable basins. Lastly, we inspect the behavior of the neural network by studying its relation to physical variables. Through the identification of its most relevant features, we are able to gain chemical insight into the process.



Keywords Enhanced sampling | Collective variables | Deep learning

Introduction

In the last decades, enhanced sampling methods have become one of the main tools of contemporary science. They have managed to extend the scope of atomistic simulations, thus allowing long time scale phenomena to be simulated. Starting from the pioneering work of Torrie and Valleau [1], a large class of such methods relies on the introduction of collective variables (CVs) defined as functions of the atomic coordinates. Once these variables are identified, to accelerate sampling a bias potential is added to the interaction potential [2]. In order to achieve this goal, the CVs are chosen so as to describe the hard-to-sample modes of the system. Not surprisingly, much effort has been devoted to the identification and improvement of useful CVs. Very recently, machine learning methods with different levels of complexity have been also used to this effect [3–8]. Here we describe a new method that is based on deep neural networks (NNs). However, before describing our approach we recall some recent findings that give a clue to what we plan to do.

The free energy landscape of physical systems can be described as made up of islands of metastability in a sea of improbable configurations. The metastable states are connected by narrow passageways that allow rare but crucial transitions from one state to another to take place. Very recently our group and others [9, 10] have developed a family of efficient CVs based only on the fluctuations that the system undergoes in the metastable basins. This defies the conventional wisdom that for a variable to be effective, it must contain explicit information about the whole reaction path or at least the transition state [3, 11].

As an example of this approach, we briefly recall how the method called harmonic linear discriminant analysis (HLDA) [9] works. One first identifies a small set of descriptors capable of discriminating between the states. The expectation value of the descriptors and their covariance matrices are computed from short unbiased runs in the two basins. With the knowledge of only these quantities, and by using a variant of the classification method that goes under the name of linear discriminant analysis (LDA) [12], one obtains CVs that are linear combinations of the input descriptors.

This simple approach has proven to be highly effective in a number of examples [13–15] but presents two limitations. At first, its linearity makes it not very effective in describing complex processes. Secondly, it relies on the identification of a small set of uncorrelated descriptors. While this is a far less demanding task than finding CVs, it requires knowledge of the system and physical intuition. This intuition can be of great help in leading to an accelerated sampling, however, it sometimes reflects more our prejudice than the actual system behavior, possibly preventing the exploration of some of the relevant transition pathways. With the use of an appropriately designed neural network, we want to lift these two limitations and construct a minimal set of CVs in order to study rare events.

Method

Once the purpose has been clarified, we describe our approach. We want to generalize the previous work of HLDA by employing a neural network and casting LDA as the objective function. This can be seen as a non-linear generalization of LDA, as proposed already in the context of computer science for classification purposes [16]. Consider a set of N data points $\mathbf{x}_1, \dots, \mathbf{x}_N$ (observations) of dimension d (local descriptors) belonging to C classes (metastable states). The covariance matrix over all these samples can be decomposed in two terms, the so-called between class and within class scatter matrices. The former takes into account the fluctuations inside the basins, and corresponds to the average of the class covariances \mathbf{S}_i :

$$\mathbf{S}_w = \frac{1}{C} \sum_i^C \mathbf{S}_i. \quad (1)$$

On the other hand, \mathbf{S}_b is defined as the covariance of the class means $\boldsymbol{\mu}_i$, and thus measures the fluctuations between classes:

$$\mathbf{S}_b = \frac{1}{C} \sum_i^C (\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T \quad (2)$$

where $\boldsymbol{\mu}$ is the average of the C class means. Fisher's criterion seeks a linear projection \mathbf{W} into a $C - 1$ dimensional space, such that the samples show a high variance between classes and a low variance within. This is achieved by maximizing Fisher's ratio:

$$\operatorname{argmax}_{\mathbf{W}} \frac{\mathbf{W}\mathbf{S}_b\mathbf{W}^T}{\mathbf{W}\mathbf{S}_w\mathbf{W}^T}. \quad (3)$$

In order to find the combination \mathbf{W} which maximizes eq. 3, one has to solve the generalized eigenvalue problem:

$$\mathbf{S}_b \mathbf{w} = \mathbf{v} \mathbf{S}_w \mathbf{w} \quad (4)$$



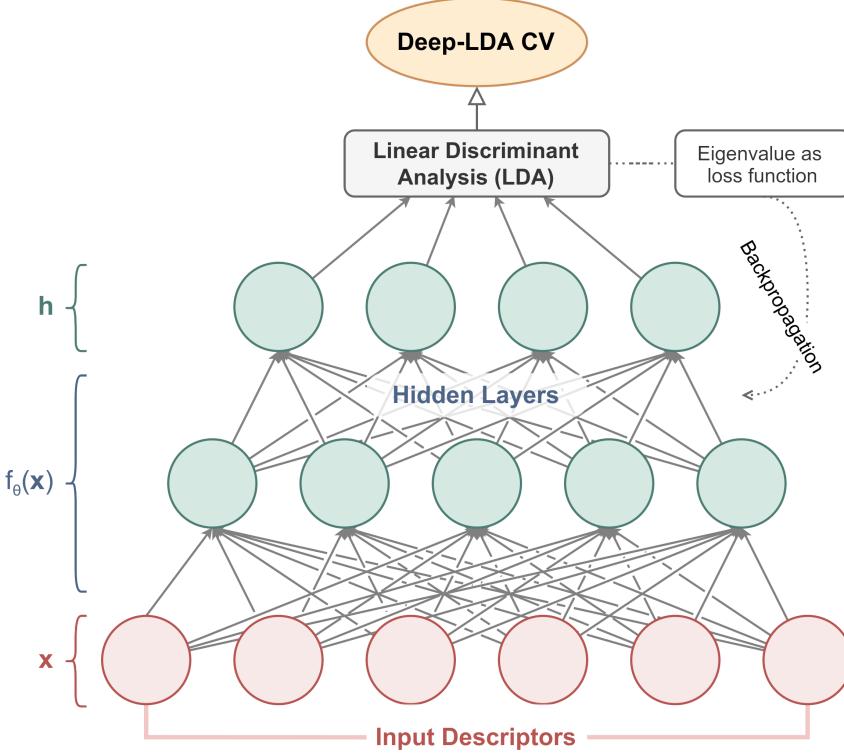


Figure 1: Scheme of the construction of the Deep-LDA CV. A set of physical descriptors are used as inputs of a feed-forward neural network. The objective is to maximize the eigenvalues of LDA on the topmost hidden representation. The corresponding eigenvectors are then used to project the last layer onto a set of CVs.

where the eigenvectors \mathbf{w} form the projection matrix \mathbf{W} and the eigenvalues \mathbf{v} quantify the separation in the corresponding directions.

We now illustrate how LDA can be used to train a neural network. We feed a number of descriptors to the NN that reduces the dimensionality of the data through a succession of continuous non-linear transformations represented by $f_\theta(\mathbf{x})$ with parameters θ (see fig. 1). We then compute LDA on the topmost hidden space representation $\mathbf{h} = f_\theta(\mathbf{x})$, and use the eigenvalues in the objective function to optimize the parameters of the NN. In the typical case of a two-classes problem, this corresponds to maximizing the single relevant eigenvalue, while for multi-states problems one can either use the smallest of the $C - 1$ eigenvalues or their sum [16]. Therefore, one increases the discriminative power of the neural network, by learning at the same time linearly separable latent features. Lastly, the hidden features are projected along the eigenvectors to obtain a minimal set of CVs. This formulation allows handling a large number of inputs, which do not need to be linearly separable and can also be correlated. We refer to this method as Deep-LDA CV.

In practice, we transform the generalized eigenvalue problem into a standard one by making first a Cholesky decomposition of $\mathbf{S}_w = \mathbf{L} \mathbf{L}^T$ [12]. By simple manipulation, we rewrite eq. 4 as:

$$\tilde{\mathbf{S}} \tilde{\mathbf{w}}_i = v_i \tilde{\mathbf{w}}_i \quad (5)$$

where $\tilde{\mathbf{S}} = \mathbf{L}^{-1} \mathbf{S}_b (\mathbf{L}^T)^{-1}$ and $\tilde{\mathbf{w}}_i = (\mathbf{L}^T)^{-1} \mathbf{w}_i$. This way, we have a symmetric eigenvalue problem which can be solved within the Pytorch library [17]. This allows training Deep-LDA networks with backpropagation and mini-batch gradient descent, provided that the batch-size is large enough to be representative of the population.

To make the LDA cost function more suitable for neural network optimization and to stabilize smaller eigenvalues, we follow [16] and regularize the within scatter matrix by adding a multiple of the identity matrix $\mathbf{S}_w + \lambda \mathbf{I}$. Since we want to use the output of the NN to enhance sampling rather than classifying, we prevent the projected states from becoming too narrow. To enforce this, we include a penalty term on the distance between the states along the CV s , defined in a smooth way by a Lorentzian function:

$$\mathcal{L}_s = -\frac{1}{1 + \left(\frac{1}{n_b n_{cv}} \sum_{b,i} s_i^2(\mathbf{x}_b) - 1 \right)^2} \quad (6)$$

where the sum goes over the elements b in the batch and the CV dimensions i . In addition, a standard L2 regularizing term is used (see S1 for the details).

After training our CV, we use it in combination with enhanced sampling methods. This could be done with any CV-based methods, such as Metadynamics [18] or Variational enhanced sampling [19, 20]. In this work, we choose to employ a recently developed variant of Metadynamics, called on-the-fly probability enhanced sampling (OPES) [21], which is based on a Gaussian modeling of the probability distribution estimate. The main benefits of OPES are the small number of free parameters, the ability to limit the exploration in the CV space to the desired process, and a faster convergence than Metadynamics. In OPES one establishes the following relation between the unbiased probability $P_n(\mathbf{s})$ and the bias potential $V_n(\mathbf{s})$:

$$V_n(\mathbf{s}) = \left(1 - \frac{1}{\gamma}\right) \frac{1}{\beta} \log(P_n(\mathbf{s}) + \epsilon) \quad (7)$$

where β is the inverse temperature, γ is the bias factor that broadens the target distribution as in well-tempered Metadynamics [22] and ϵ is a regularization term. The probability estimate is computed on-the-fly using a weighted kernel density estimation:

$$P_n(\mathbf{s}) = \frac{\sum_k^n w_k G(\mathbf{s}, \mathbf{s}_k)}{\sum_k^n w_k} \quad (8)$$

where the weights $w_k = e^{\beta V_{k-1}(\mathbf{s}_k)}$ are connected to the value of the bias potential. Once converged, the free energy surface (FES) can be computed as:

$$F_n(\mathbf{s}) = -\frac{1}{\beta} \log P_n(\mathbf{s}) \quad (9)$$

and simple reweighting gives access to all the static properties of interest. We refer the reader to [21] for further details. Before proceeding to the applications, we summarize the method for clarity:

1. Run short unbiased MD runs in the metastable states and compute the descriptors
2. Construct a CV by training a NN with LDA as the objective function
3. Use the Deep-LDA CV to enhance the sampling and obtain the FES

Alanine dipeptide

The first example that we choose is alanine dipeptide (Ala2), a small molecule often used to benchmark sampling methods. Ala2 has two metastable states that are well described using the pair of Ramachandran angles ϕ and ψ , which forms a nearly ideal set of CVs. For this reason, the majority of enhanced methods is effective when the angles' fluctuations are enhanced. In order to test Deep-LDA in a less than ideal situation, we opted to use as input all the 45 distances between heavy atoms. This choice makes Ala2 a non-trivial example since the physical variables ϕ and ψ cannot be represented with a linear expression of the interatomic distances and the two states are not linearly separable. We run short unbiased trajectories in the two basins and use this data to train a Deep-LDA CV. The model is then imported in the open-source plug-in PLUMED [23] with a custom Pytorch interface and used in OPES. Further details are reported in S2 and S3.

Enhancing the fluctuations of the Deep-LDA CV leads to a very diffusive behavior between the two states, which is comparable to the simulations where the pair of Ramachandran angles is biased (see fig. S3.1). This means that Deep-LDA manages to reproduce in an implicit way the two dihedrals which represent the slowest degrees of freedom.

To illustrate this point further, we present in fig. 2 the isolines of the Deep-LDA CV on top of the FES, both projected onto ϕ and ψ . As there is no direct correspondence between the dihedral angles and the CV s , the isolines have been computed from the conditional probability $p(s | \phi, \psi)$ in the statistical ensemble generated by an OPES calculation with an uniform target distribution in the (ϕ, ψ) space. When the Deep-LDA fluctuations are enhanced by OPES, the system is driven along the direction perpendicular to the isolines of the CV. It is remarkable how this direction is correlated with the path that connects the two basins in the dihedral's space, managing to trace the underlying free energy landscape (see also fig. S3.2).

We note here that, since Ala2 is a very small molecule, we chose to include in the descriptors set only the distances between heavy atoms. In principle, the Deep-LDA method can be applied to any set of descriptors, including but not limited to angles, dihedrals or hydrogen bonds contacts.

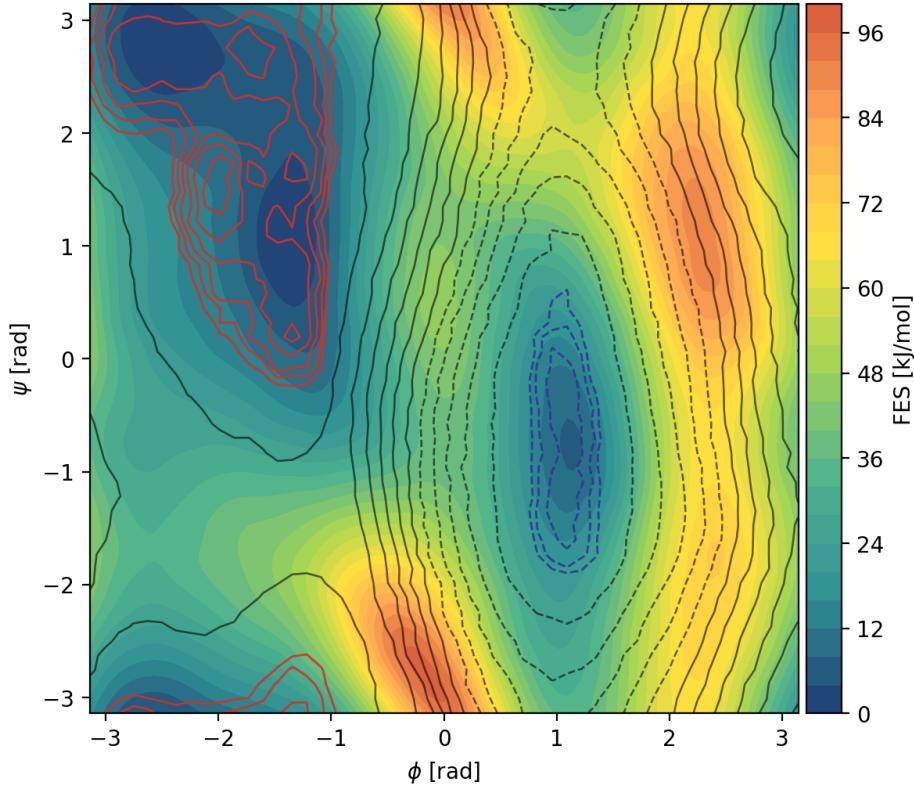


Figure 2: Free energy surface of Ala2 as a function of the Ramachandran angles, together with the isolines of the Deep-LDA CV, computed from a simulation with a uniform target distribution in this space. The red and blue lines correspond to the regions of the two metastable states used for training, and they are spaced by 0.01. The black ones are extrapolated by the network, with a much larger spacing of 0.2. The isolines are solid or dashed depending on whether they are above or below zero.

Aldol reaction

The second example is the aldol reaction between vinyl alcohol and formaldehyde. This reaction presents several pathways and products [24]. Here, for simplicity, we focus on the most probable one, that has a barrier of ~ 150 kJ/mol as estimated in static calculations. This represents a challenging test since it is not obvious whether the reaction's concerted mechanism can be captured with information coming exclusively from unbiased simulations of reactants and products. Once again, in the spirit of performing the computation in a blind manner, we build the descriptor set out of all the interatomic distances. We rely on the ability of Deep-LDA to combine them in a meaningful and optimal way and to highlight the most relevant ones. Previous experience [15] has shown that a good input set for chemical reactions can be obtained from the distances r_{ij} using the following contact function:

$$c_{ij}(r) = \frac{1 - \left(\frac{r}{\sigma_{ij}}\right)^n}{1 - \left(\frac{r}{\sigma_{ij}}\right)^m} \quad (10)$$

where σ_{ij} are typical bonding lengths between atoms of species i and j . Thus, we express the fact that following the formation and breaking of chemical bonds is key to capturing a chemical process. In this way we still maintain a blind generality while injecting a minimal amount of chemical knowledge. As in the case of modeling interatomic potentials with NNs, the use of chemically informed descriptors allows the NN to learn more easily from the data. See S4 for the computational details.

As before, we train the Deep-LDA network on unbiased simulations of the two states and use this CV in combination with OPES to enhance the process. One representative trajectory is reported in fig. 3a, which illustrates how the system is reversibly driven from reactants (red) to products (blue).

We then focus our attention to the interpretability of the model, by analyzing which features are most important in the Deep-LDA CV. The ranking is defined by summing the modulus of the weights between the input and the first layer, multiplied by the standard deviation of the inputs in the training set. Another option is to calculate the derivatives of the output of the NN with respect to each input, averaged over all inputs in the training set. We found that these two ways of estimating the feature relevance produce similar results. As shown in fig. 3b, there are three relevant descriptors, separated by a large gap from the following ones. These are the C1-C2 contact where the bond is formed together with the O1-H and O2-H contacts that characterize the proton transfer. This result has been obtained without any *a priori* information on the reaction pathway and it is remarkably in agreement with chemical intuition.

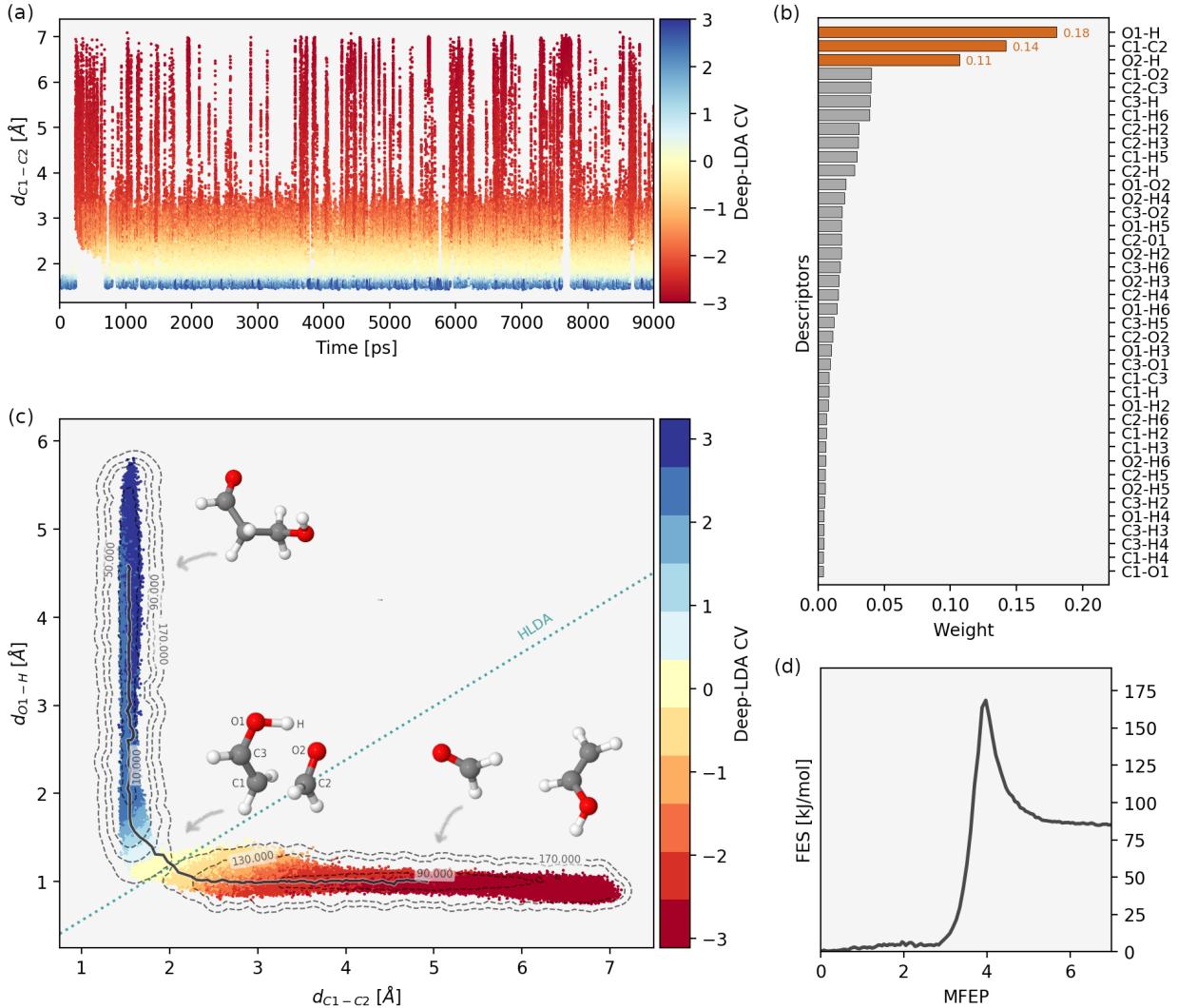


Figure 3: Results of the OPES simulation when enhancing the Deep-LDA CV. (a) Evolution of the carbon-carbon distance between vinyl alcohol and formaldehyde as a function of simulation time. The points are colored according to the Deep-LDA CV. (b) Features ranking according to the magnitude of the weights of the first layer. They are normalized such that the sum is equal to one. The first three distances, separated by a gap from the following ones, are colored in orange. (c) The points correspond to the visited configurations in the plane of the C1-C2 and O1-H distances. As in panel (a), the color is the Deep-LDA CV. The gray dashed lines correspond to the isolines of the free energy surface projected onto this space, while the black solid one corresponds to the minimum free energy path. The teal dotted line is the boundary decision of HLDA. We also reported a few snapshots of the representative structures of the two metastable states and the region between them. (d) Free energy surface projected along the minimum free energy path, computed using the nudged elastic band method [25].

The FES projected on a pair of these chemically relevant variables presents an easily understandable structure (fig. 3c and S4.2). The two basins are clearly separated and the Deep-LDA CV shows different values in the two basins. As in the Ala2 example, the isolines of Deep-LDA resemble quite closely the ones of the FES. This implies that the direction where the system is driven is correlated with the MFEP, thus acting as a committor-like collective variable. Enhancing the fluctuations along this CV accelerates the sampling along the entire path between the metastable states. This features improves on HLDA that always drives the system in the direction perpendicular to the boundary line drawn in fig. 3c (see also fig. S4.3).

The progress of this work is the realization of a CV that accompanies the evolution of the system along the chemical pathway. We believe this approach can be crucially useful in more complex cases where there are several reaction pathways or multiple reactions and the identification of relevant descriptors is not trivial.

Conclusions

Deep-LDA is able to compress the information from the metastable states into collective variables. Thanks to the LDA objective function, the distribution of the CV is characterized by a high variance between classes and a low variance within the same class. The non-linearity provided by neural networks allows following the fluctuations between the states, effectively drawing a path in the descriptors space. In principle, the Deep-LDA CV can be used in combination with any CV-based enhanced sampling method. We want to remind the reader that our purpose here is not to find the ideal CV, rather to build a variable which is good enough to promote transitions between the metastable states, given only a limited amount of information. This CV could be further refined in order to closely follow the reaction pathway, either by including weighted data from biased simulations or by combining it with methods designed to extract the slowest relaxation modes [26–30]. Another result is that the Deep-LDA features ranking can be used to identify the relevant descriptors in a data-driven way, as well as filtering them. For all these reasons we believe that our method could be of help in studying a large variety of rare events including but not limited to chemical reactions, nucleation events and ligand-binding processes.

Acknowledgments

The authors thank Manyi Yang for providing the inputs of the aldol reaction and Sandro Bottaro, Riccardo Capelli, Emanuele Grifoni, Michele Invernizzi and GiovanniMaria Piccini for useful discussions. This research was supported by the NCCR MARVEL, funded by the Swiss National Science Foundation, the Swiss National Science Foundation Grant No. 200021_169429/1 and the European Union Grant No. ERC-2014-AdG-670227/VARMET. The calculations were carried out on the Euler cluster of ETH Zurich.

SUPPORTING INFORMATION

S1 Neural network training

The Deep-LDA model is trained using the PyTorch [17] library. We report here the parameters used for the training of the neural network. We first apply a regularization to the within class scatter matrix in the form of $\mathbf{S}_w + \lambda \mathbf{I}$, with $\lambda = 0.05$.

The following loss function is used:

$$\mathcal{L} = \epsilon_0 + \alpha \mathcal{L}_s + \gamma \sum_i |\theta_i|^2 \quad (\text{S1})$$

where ϵ_0 is the largest eigenvalue of eq. 4, \mathcal{L}_s is the Lorentzian loss function on the output given by eq. 6, and the third term is an L2 regularization over the weights θ_i of the network, with $\gamma = 10^{-5}$. We found that the intensity α of the penalty function \mathcal{L}_s is connected to the value of the regularizing term λ , and a robust behaviour is obtained when their product is kept constant. In fact, α affects the numerator of Fisher's ratio (eq. 4), while the latter only the denominator. Thereby we use $\alpha = 2/\lambda$. We use the rectified linear unit as activation function. The training data, which consists of 10000 configurations, is divided into batches of size 2000, and the model is optimized using ADAM [31] with a learning rate of 10^{-4} , until the loss function converges. In order to prevent overfitting, a small set of configurations from different simulations is used as a validation set, and the early stopping technique is used where necessary. Once the model has been trained, the LDA coefficients are obtained using all the training set and the model is exported.

S2 PLUMED interface with PyTorch

A development version of the open-source plug-in PLUMED [23] is used. The first task is the definition of the descriptors, which form the input for the Deep-LDA network. The NN architecture and parameters are loaded using a custom interface via LibTorch C++ APIs. Finally, the Deep-LDA collective variable is used in combination with the OPES method to enhance the sampling. We found that using the transformation $x' = x + x^3$ for the Deep-LDA CV helps OPES dealing with the small standard deviation which characterizes the two basins.

A typical input script for PLUMED is the following:

```
# --- (1) DESCRIPTORS ---
d1: DISTANCE ATOMS=1,2
d2: DISTANCE ATOMS=1,3
...
dN: DISTANCE ATOMS=17,19

# --- (2) DEEP-LDA CV ---
deep: PYTORCH_MODEL MODEL=model.pt ARG=d1,d2,...,dN
deep_cube: MATHEVAL ARG=deep.0 FUNC=x+cube(x) PERIODIC=NO

# --- (3) BIAS POTENTIAL ---
opes: OPES MODE=CONVERGE ARG=deep_cube PACE=500 SIGMA=0.05 BARRIER=40
```

S3 Alanine dipeptide

The alanine dipeptide (Ala2) simulations are carried out using GROMACS [32] patched with PLUMED. We use the Amber99-SB force field [33] with a time step of 2 fs. The NVT ensemble is sampled using the velocity rescaling thermostat [34] with a temperature of 300K. For OPES we use SIGMA=0.05 and BARRIER=40 kJ/mol. The neural network used to model the Deep-LDA CV has 3 hidden layers with 30, 15, 5 neurons per layers. The distances are first standardized to take values between -1 and 1.

In the top panel of fig. S3.1 we report the trajectory of a simulation using Deep-LDA, compared with the case in which the two Ramachandran angles are used. As it can be seen from the transition rate, despite the fact that only scalar distances are used as inputs for Deep-LDA, the NN manages to build a non-linear combination of them which is able to distinguish and to drive the transition between the states, thus mimicking the effect of the diedhral angles.

In fig. S3.2 we report the probability distribution of the Deep-LDA CV as a function of the two angles, $p(s|\phi, \psi)$. Next to it, we report the free energy surface obtained from the OPES simulation via a reweighting technique [21]:

$$P(s) = \frac{\langle \delta[s - s(\mathbf{R})] e^{\beta V(s)} \rangle_V}{\langle e^{\beta V(s)} \rangle_V} \quad (\text{S2})$$

By comparing S3.2 and S3.3, we notice that Deep-LDA effectively reproduces the topology of the free energy landscape of Ala2, while using only short unbiased simulations. It is encouraging that the Deep-LDA isolines of fig. S3.2 are characterized by a strong correlation with the ϕ angle, which is the slowest degree of freedom.

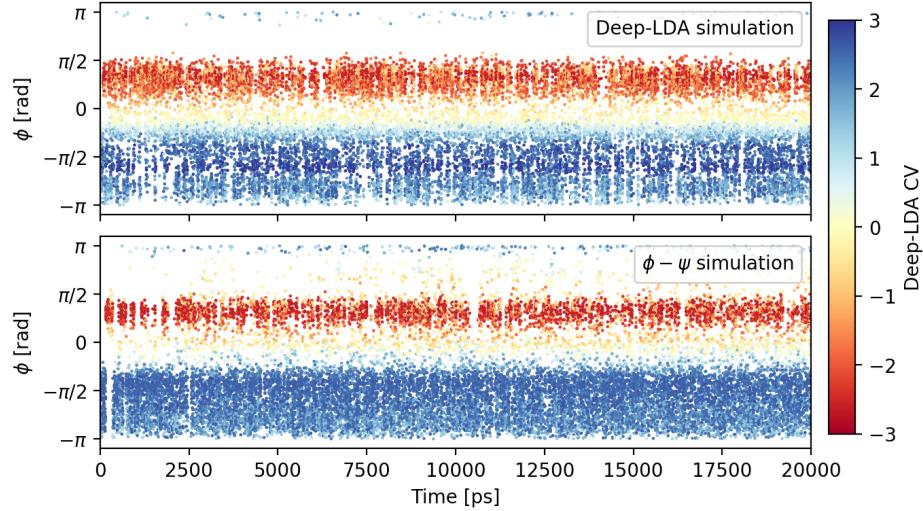


Figure S3.1: Time evolution of the ϕ angle in the Deep-LDA OPES simulation (top) and the optimal case when the two dihedral angles are used to define the bias potential (bottom).

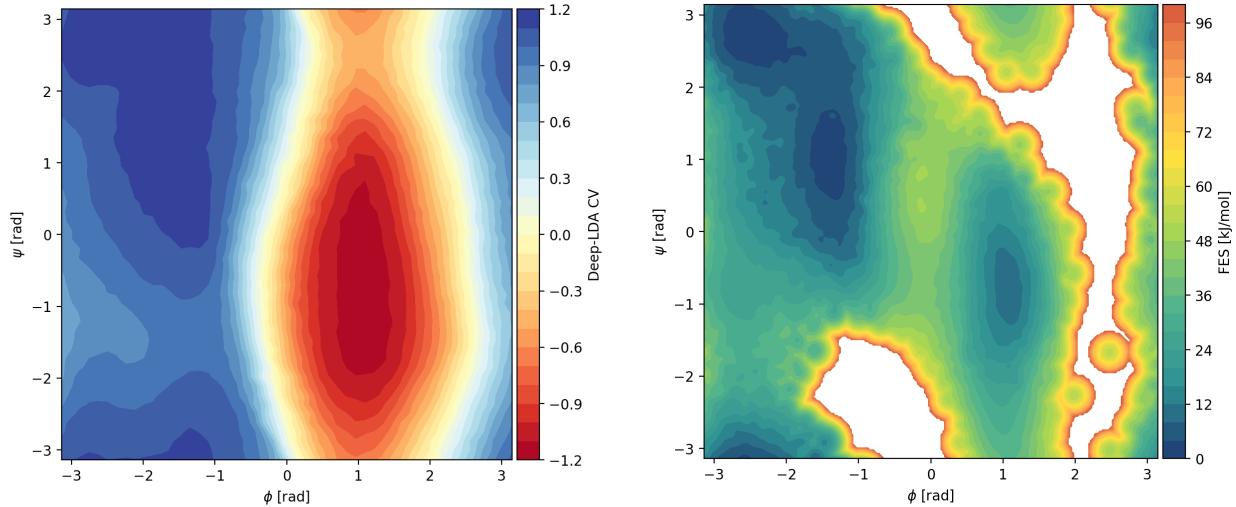


Figure S3.2: Ramachandran plot with the isolines of Deep-LDA.

Figure S3.3: Free energy profile of Ala2 along the two dihedral angles.

S4 Aldol reaction

The software package CP2K [35] is used to carry out the simulations of the chemical reaction, at the PM6 semi-empirical level. We used an integration step of 0.5 fs, employing the same thermostat as in S3 with a time constant of 100 fs. In fig. S4.1 we report a snapshot of the aldol reaction with the labeling of the atoms used in the paper. The contacts between the atoms are computed using PLUMED, with the parameter σ_{ij} of the switching function of eq. 10 being equal to 1.7 Å for C-C, 1.6 Å for O-O and C-O, and 1.2 Å for C-H and O-H species. The exponents of the contact function are chosen as $n = 6$ and $m = 8$ to enforce a smooth behavior over a wide range of distances. The full set is made of 40 contacts. A neural network with four hidden layers and (30,20,10,5) nodes per layer is used. For OPES we use SIGMA=0.05 and BARRIER=160 kJ/mol.

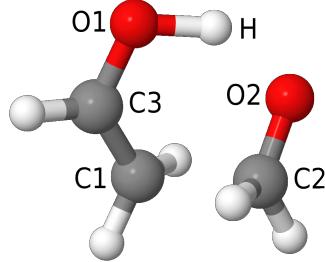


Figure S4.1: Snapshot of the aldol reaction with labels on the relevant atoms involved in the process. The C1-C2 bond has to be formed, with a simultaneous proton transfer from O1 to O2.

We report in fig. S4.2 the free energy profiles obtained via reweighting of the Deep-LDA CV in the space of the three chemical distances identified by the features ranking.

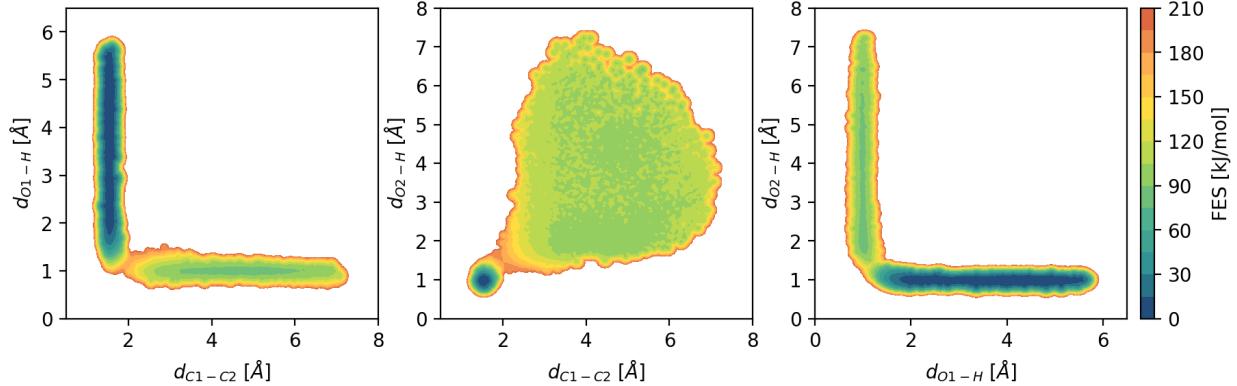


Figure S4.2: Free energy surface projected along combinations of the three chemical distances identified as the most important from the Deep-LDA ranking.

We note that HLDA cannot be applied in the same blind way to the whole set of contacts, because they are strongly correlated and a linear combination is not effective in such a high-dimensional space. However, once the relevant descriptors are identified with the Deep-LDA features ranking, we can use HLDA on the reduced space to build a simpler and more physically interpretable CV. We apply HLDA on the C1-C2, O1-H, O2-H contacts and use it to enhance the sampling as a comparison with Deep-LDA (fig. S4.3). We see that also HLDA is able to promote transitions, even though it explores the two metastable states in a less efficient way, as it always drives the system in the same linear direction.

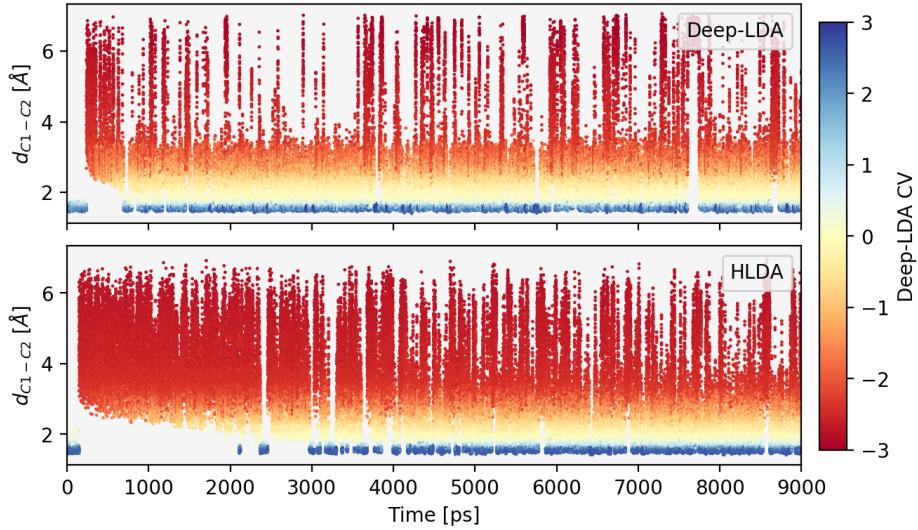


Figure S4.3: Time evolution of the C1-C2 distance in simulations in which Deep-LDA (top), and an HLDA combination of the three distances C-C, O1-H, O2-H (bottom) is used as CV.

References

- [1] G. Torrie and J. Valleau. “Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling”. In: *J. Comput. Phys.* 23.2 (1977), pp. 187–199.
- [2] O. Valsson, P. Tiwary, and M. Parrinello. “Enhancing Important Fluctuations: Rare Events and Metadynamics from a Conceptual Viewpoint”. In: *Annu. Rev. Phys. Chem.* 67.1 (2016), pp. 159–184.
- [3] A. Ma and A. R. Dinner. “Automatic Method for Identifying Reaction Coordinates in Complex Systems ”. In: *J. Phys. Chem. B* 109.14 (2005), pp. 6769–6779.
- [4] J. M. L. Ribeiro, P. Bravo, Y. Wang, and P. Tiwary. “Reweighted autoencoded variational Bayes for enhanced sampling (RAVE)”. In: *J. Chem. Phys.* 149.7 (2018), p. 072301.
- [5] W. Chen and A. L. Ferguson. “Molecular enhanced sampling with autoencoders: On-the-fly collective variable discovery and accelerated free energy landscape exploration”. In: *J. Comput. Chem.* 39.25 (2018), pp. 2079–2102.
- [6] M. Schöberl, N. Zabaras, and P.-S. Koutsourelakis. “Predictive collective variable discovery with deep Bayesian models”. In: *J. Chem. Phys.* 150.2 (2019), p. 024109.
- [7] J. Rogal, E. Schneider, and M. E. Tuckerman. “Neural-Network-Based Path Collective Variables for Enhanced Sampling of Phase Transformations”. In: *Phys. Rev. Lett.* 123.24 (2019), p. 245701.
- [8] P. Ravindra, Z. Smith, and P. Tiwary. “Automatic mutual information noise omission (AMINO): generating order parameters for molecular systems”. In: *Mol. Syst. Des. Eng.* 5.1 (2020), pp. 339–348.
- [9] D. Mendels, G. Piccini, and M. Parrinello. “Collective Variables from Local Fluctuations”. In: *J. Phys. Chem. Lett.* 9.11 (2018), pp. 2776–2781.
- [10] M. M. Sultan and V. S. Pande. “Automated design of collective variables using supervised machine learning”. In: *J. Chem. Phys.* 149.9 (2018), p. 094106.
- [11] B. Peters. *Reaction Rate Theory and Rare Events*. Elsevier, 2017, pp. 1–619.
- [12] M. Welling and Max Welling. “Fisher linear discriminant analysis”. In: *Dep. Comput. Sci. Univ. Toronto* (2005).
- [13] G. Piccini, D. Mendels, and M. Parrinello. “Metadynamics with Discriminants: A Tool for Understanding Chemistry”. In: *J. Chem. Theory Comput.* 14.10 (2018), pp. 5040–5044.
- [14] D. Mendels, G. Piccini, Z. F. Brotzakis, Y. I. Yang, and M. Parrinello. “Folding a small protein using harmonic linear discriminant analysis”. In: *J. Chem. Phys.* 149.19 (2018), p. 194113.
- [15] V. Rizzi, D. Mendels, E. Sicilia, and M. Parrinello. “Blind Search for Complex Chemical Pathways Using Harmonic Linear Discriminant Analysis”. In: *J. Chem. Theory Comput.* 15.8 (2019), pp. 4507–4515.
- [16] M. Dorfer, R. Kelz, and G. Widmer. “Deep linear discriminant analysis”. In: *4th Int. Conf. Learn. Represent. ICLR*. 2016.
- [17] A. Paszke et al. “Automatic differentiation in PyTorch”. In: *Adv. Neural Inf. Process. Syst.* 32. 2019, pp. 8024–8035.
- [18] A. Laio and M. Parrinello. “Escaping free-energy minima”. In: *Proc. Natl. Acad. Sci.* 99.20 (2002), pp. 12562–12566.
- [19] O. Valsson and M. Parrinello. “Variational Approach to Enhanced Sampling and Free Energy Calculations”. In: *Phys. Rev. Lett.* 113.9 (2014), p. 090601.
- [20] L. Bonati, Y.-Y. Zhang, and M. Parrinello. “Neural networks-based variationally enhanced sampling”. In: *Proc. Natl. Acad. Sci.* 116.36 (2019), pp. 17641–17647.
- [21] M. Invernizzi and M. Parrinello. “Rethinking Metadynamics: from bias potential to probability distribution”. In: *arXiv, id:1909.07250* (2019).
- [22] A. Barducci, G. Bussi, and M. Parrinello. “Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method”. In: *Phys. Rev. Lett.* 100.2 (2008), p. 020603.
- [23] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi. “PLUMED 2: New feathers for an old bird”. In: *Comput. Phys. Commun.* 185.2 (2014), pp. 604–613.
- [24] M. Yang, J. Zou, G. Wang, and S. Li. “Automatic Reaction Pathway Search via Combined Molecular Dynamics and Coordinate Driving Method”. In: *J. Phys. Chem. A* 121.6 (2017), pp. 1351–1361.
- [25] H. Jonsson, G. Mills, and K. W. Jacobsen. “Nudged elastic band method for finding minimum energy paths of transitions”. In: *Class. Quantum Dyn. Condens. Phase Simulations*. World Scientific, 1998, pp. 385–404.
- [26] J. McCarty and M. Parrinello. “A variational conformational dynamics approach to the selection of collective variables in metadynamics”. In: *J. Chem. Phys.* 147.20 (2017), p. 204109.
- [27] P. Tiwary and B. J. Berne. “Spectral gap optimization of order parameters for sampling complex molecular systems”. In: *Proc. Natl. Acad. Sci. U. S. A.* 113.11 (2016), pp. 2839–2844.

- [28] W. Chen, H. Sidky, and A. L. Ferguson. “Nonlinear discovery of slow molecular modes using state-free reversible VAMPnets”. In: *J. Chem. Phys.* 150.21 (2019), p. 214114.
- [29] C. Wehmeyer and F. Noé. “Time-lagged autoencoders: Deep learning of slow collective variables for molecular kinetics”. In: *J. Chem. Phys.* 148.24 (2018), p. 241703.
- [30] C. X. Hernández, H. K. Wayment-Steele, M. M. Sultan, B. E. Husic, and V. S. Pande. “Variational encoding of complex dynamics”. In: *Phys. Rev. E* 97.6 (2018), p. 062412.
- [31] D. P. Kingma and J. Ba. “Adam: A Method for Stochastic Optimization”. In: *3rd Int. Conf. Learn. Represent. ICLR*. 2014.
- [32] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. “GROMACS: Fast, flexible, and free”. In: *J. Comput. Chem.* 26.16 (2005), pp. 1701–1718.
- [33] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling. “Comparison of multiple Amber force fields and development of improved protein backbone parameters”. In: *Proteins Struct. Funct. Bioinforma.* 65.3 (2006), pp. 712–725.
- [34] G. Bussi, D. Donadio, and M. Parrinello. “Canonical sampling through velocity rescaling”. In: *J. Chem. Phys.* 126.1 (2007), p. 014101.
- [35] J. Hutter, M. Iannuzzi, F. Schiffmann, and J. VandeVondele. “Cp2k: atomistic simulations of condensed matter systems”. In: *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 4.1 (2014), pp. 15–25.